

An Institutional Perspective on Genres: Generic Subtitles in German Literature from 1500-2020

Benjamin Gittel

Benjamin Gittel, University of Göttingen

Peer-Reviewer: Gerhard Lauer, Matthew Wilkens

Dataverse DOI: 10.7910/DVN/XZAXTE

ABSTRACT

Using a custom-designed database of 388,000 first editions of German Literature this paper investigates the long-term development of genre-indicating subtitles over more than 500 years of literary history. This approach adds a social-institutional perspective to recent work in the field of genre theory, and is a first step towards combining historical testimony, i.e. historical actors' classifications, and textual features in a single model. Starting from the fundamental question of how many books have generic subtitles, the paper analyses the use of the most common genre labels, the relation between generic subtitles and genre production, periods of the permanent presence of generic terms (institutional cycles) and periods of generic differentiation. It identifies recurrent patterns in the development of generic subtitles using K-Means-Clustering and Dynamic Time Warping (DTW) and sheds light on literature's changing relation to history and truth, thereby underpinning recent theoretical work on the practices of poetic invention.

It would be worthwhile to distinguish two aspects of research on literary genres.¹ First there is research on *text types* as generalizations over textual features, and second, *genres* as generalizations over individuals' attitudes to texts.² Although the latter aspect seems crucial for a perspective that does justice to the social dimensions of literature and its history, recent research in the Digital Humanities has most frequently focused on literary texts as text types.³ Descriptive models aim at determining the most distinctive features of one group of texts versus another, while predictive models have tried to reproduce or even predict human classifications of literary texts,⁴ usually working with the classifications of librarians or contemporary researchers. Although predictive modeling does not aim at generalizations over the textual features that literary scholars have in mind when they define text types (e.g. "All texts that start with 'once upon a time'") — arguably, it does not aim at definitions at all —⁵ it clearly relies on textual features or, more precisely, probabilistic combinations of textual features.

As impressive as the successes of this approach have been, however, it represents just one perspective on genre that ought to be supplemented by a social-institutional perspective. Genre concepts are and have been used by historical actors to orient and position themselves in the literary field and are accordingly linked to certain expectations (e.g. “All texts from which the reader expects a moral”). Such expectations, which are often linked to certain terms (e.g. “fable”), can become established in certain periods of time and disappear in later periods of time. In other words, a genre can become institutionalized or deinstitutionalized.⁶ Yet this sort of social-institutional perspective on genres has rarely been pursued in the digital humanities,⁷ in part because computational approaches are generally conceived as complementary to classical genre research: “Text analysis,” argues Ted Underwood and the NovelTM Research Group, “may not provide a complete picture of the history of genre, but it does give us a second reference point - a touchstone that we can use to unpack historical assertions about genre, or compare them to each other.”⁸ Complementing this point of view, I would argue here for the use of statistical computational methods with regards to the problem of institutionalization, which lies at the heart of historically oriented genre theories in literary studies. The following analysis of the occurrence of genre-indicating concepts in literary works’ titles or subtitles over roughly 500 years of literary history is a first step on this path.

Before presenting my dataset (section 1), methods (section 2), and results (section 3), I want to point out more in detail what I mean by the term “institutionalization” and how this article will tackle the problem of genre institutionalization. Obviously, an account of the institutionalization of a genre can take various forms. Against the backdrop of the distinction that has been introduced in the first paragraph, one can think of the institutionalization process as the process by which — in addition to the instantiation of a text type *T* in a certain period of time — additional conditions are met that include a certain attitude of people towards texts of the type *T*. There are at least three candidates for these additional conditions that must be met for a genre to be institutionalized or, in more simpler words, to exist in a certain period (in a certain community): (1) feature expectations, (2) genre awareness, and (3) rules for dealing with texts of type *T*.⁹ Fairy tales, for example, arguably fulfil all three conditions in our days. When we read a text in which dragons, elves and fairies appear, we expect the appearance of the wizard with a long white beard (feature expectations). Furthermore, we are aware that the genre “fairy tale” exists (genre awareness). Third,

when reading fairy tales, we accept that dragons are able to spit fire (it may not even have to be said) and we are usually not interested in the spatio-temporal location of the story (rules for dealing with a particular text type). Depending on which of these three combinable criteria a genre concept includes, differently sophisticated concepts of genre institutionalization result.

The relevance of generic concepts' occurrence in the titles or subtitles of literary works for an institutional approach results directly from the account of genre institutionalization just introduced, which has been elaborated elsewhere.¹⁰ For the approach chosen in this article, feature expectations and genre awareness are the important conditions:¹¹ Only if the authors, publishers, or editors of literary works believe that readers will associate certain expectations with a particular genre concept, will they use it with any frequency as a title or subtitle.¹² Therefore, generic concepts' occurrence in the titles or subtitles of literary works can be regarded as indicators for the institutionalization or deinstitutionalization of a literary genre. Please note that this consideration applies irrespective of the emergence of a literary market in the modern sense and irrespective of a possible connection between a genre's production on the one hand and the occurrence of the respective genre term in subtitles on the other hand (a connection argued for in section 3.3).

One final note before getting started: Since there is very few quantitative research on the subject of this paper, at least for German literature,¹³ it may be sometimes difficult to classify the results. I have tried to point out links to existing qualitative research wherever it seemed useful and where my limited knowledge on the broad investigation period allowed it. My expectation was not to prove or disprove any particular theory or view, but the conviction that the approach chosen here will prove fruitful and offer a variety of different connecting points to past and future research.

1. Dataset

The database of German Literature needed for this article's analysis had to fulfil four criteria:

- be as comprehensive as possible;
- contain only literature in German without translations into German;
- comprise only first editions; and

- contain no duplicates.

These criteria resulted directly from the intention to analyse the frequency of generic terms in titles and subtitles. Literature in other languages would distort the analysis, as would translations from other languages, because foreign genre-related subtitles are often not easily translatable into German and are therefore frequently changed or omitted. Later editions would also distort the analysis, insofar as they often preserve subtitles from original publications and might thus virtually prolong the existence of old-fashioned genre terms.

Since there was no existing database of German literature that would even marginally comply with these criteria, it had to be developed for this article.¹⁴ In order to do so, I worked with one of the biggest library networks in Germany, the GBV (Gemeinsamer Bibliotheksverbund), which encompasses more than 430 libraries, including the national library in Berlin (The Staatsbibliothek zu Berlin) and the German National Library DNB (Deutsche Nationalbibliothek in Leipzig/Frankfurt). GBV's data, as provided in MARC21 format, comprised approximately 68 million records (around 120 GB), which could only be filtered according to this article's criteria with invaluable information provided by various GBV and DNB employees.¹⁵ The filtering proved relatively complex: the data originated from different libraries and was created during different time periods and regions, meaning that different classification systems and codes were used for so-called subject groups (*Sachgruppen*). The following filtering was carried out in Python (see table 1).¹⁶

Step	Records	Description
1. Starting point	68,366,626	Total GBV records
2. Filtering of books	45,194,326	Books (no e-books, no periodicals)
3. Filtering of German Literature	740,053	German Literature (<i>Schöne Literatur</i>) without translations
4. Filtering of first editions	472,502	First edition German Literature without translations
5. Deleting of duplicates	388,127	First edition German Literature without translations and without duplicates

Table 1: Filtering of GBV-library-data

Since only around 60% of all of the records in the GBV library data had a language code, automatic language detection based on titles and subtitles was performed for the records without a language code, using “detectlanguage” from Google.¹⁷ The filtering out of later editions took into account numerated editions, as well as standardized references to “supplemented edition” (*erg. Aufl.*), “extended edition” (*erw. Aufl.*), or “folk edition” (*Volksausgabe*). The duplicate search was based on author name and title (without subtitles, which may have changed in later editions), with the earliest published work retained.

In this way, a database of first edition German-language literature was created, covering 520 years of literary history (see figure 1). The database is made available to the scientific community for further research.¹⁸



Figure 1: Books per year

Apart from a slump in production during to the Second World War, the year 1800 stands out as an outlier. This is probably due to the practice of dating books published in the years around 1800 either back or forward to align with the

centennial mark. The drop at the end of the graph is due to the incompleteness of available data, especially for 2019 and 2020.¹⁹

2. Methods

Analyzing the data was essentially a matter of determining the frequencies of genre terms in titles or subtitles per year and relating these frequencies to literary production.²⁰ It proceeded in two steps: recognition of genre terms and basic descriptive statistics.

2.1 Recognition of genre terms

A distinction was made between three different types of genre terms' occurrences:

- (a) total genre appearances (e.g. the word “novel” (*roman*) in “romance novel” (*Liebesroman*), “historical novel” (*historischer Roman*), or just “novel” (*Roman*), but excluding “romantic” (*romantisch*));
- (b) compound genre terms (e.g. “romance novel” (*Liebesroman*));
- (c) phrasal genre references in the form adjective + genre (e.g. “historical novel” (*historischer Roman*));

These search patterns were elaborated through a list of 73 basic non-compound genre terms like novel (*Roman*), story (*Geschichte*), and poem (*Gedicht*) (see Annex 1).²¹ In order to effect the categorization, regular expressions in Python and the CNN-based Part-of-Speech-Tagger (“de_core_news_sm”, version 2.3.0) from the package “spaCy” were used.²²

2.2 Basic descriptive statistics

Each occurrence of a genre term in titles or subtitles was counted per year. In addition, the number of titles in which at least one basic genre term occurred was counted (search pattern “total genre appearances”). Statistics on clearly identical genre terms, which had been created by different spellings or abbreviations (for example, “hist. Roman”), were merged, i.e. the counts were summated. These absolute values were then converted into relative values by relating the counts to literary production in the year in question. This was an important step towards obtaining meaningful and comparable values, since literary production – due to

technical, historical and book market developments – has varied greatly over 500 years (see figure 1, p. 5).

2.3 Determining the most persistent genres

The full list of compound genre terms comprised 25,064 items (10,876 compound terms + 14,188 phrasal references), many of them appearing only once. In order to filter this list for later analysis, I determined the most persistent genres within the list. In a first step, I calculated a two-year rolling average in pandas (Version 1.0.5) in order to compensate for minor fluctuations in the data.²³ Second, I determined all of the cycles (time periods) in which a genre term was present for at least 10 consecutive years (hereafter: “institutional cycles”). Third, I created a list of genres that had at least one institutional cycle. The resulting list of 371 genres has been sorted out semi-automatically to exclude non-genre terms (e.g. “selected poems,” “German poems,” “most beautiful stories”), non-intended matches (e.g. *czernowitz*, a falsely recognized composition with the genre term *witz* (joke), or obviously non-literary genres (e.g. “studbook” (*Stammbuch*)). This resulted in a list of 279 genres (see appendix A2).

This list served as the basis for (a) the visualization of the institutional cycles of the most persistent subgenres (see section 3.4) and (b) the K-Means Clusters using Dynamic Time Warping (see section 2.4). To visualize the institutional cycles of the most persistent subgenres, I used the information gathered during step two of the process outlined above. Figure 8 (p. 16) shows the institutional cycles of the 40 most persistent genre labels.

I also determined the time periods in which the largest number of genres “fought” for institutionalization, that is, in the periods when many new genre labels are repeatedly employed. To do so, I used the same multi-step procedure outlined above, with one major difference. In this case, a three-year rolling average was calculated (minimum number of observations = 1), which yielded a list of 581 genres (see section 3.5). For each of the most persistent genres the starting point of its *first* institutional cycle was determined. Figure 9 (p. 18) visualizes the counts of starting points per year.

2.4 K-Means Clustering using Dynamic Time Warping (DTW)

In order to identify patterns in 279 time series, each corresponding to a persistent genre term, I used K-Means-Clustering on the basis of Dynamic Time Warping (DTW) as implemented in Python's "tslearn" (Version 0.4.1).²⁴ DTW is used to collect time series of similar shapes. It is especially suitable for the data in this article, insofar as it does not presuppose synchronized time series, meaning that it can find similar shapes even if they are offset to each other.²⁵ In practice, DTW outperforms its classical rival Euclidean Distancing on most data sets.²⁶

To prepare the data for the clustering, leading zeros and NA's were removed, and then `TimeSeriesScalerMeanVariance` and `TimeSeriesResampler` (both from the `tslearn`-package) were applied. The time series were scaled so that each output time series had zero mean and unit variance. They were resampled to a target size of 50 to limit computation time. The K-Means-Clustering was achieved with the "TimeSeriesKMeans" function, which allows soft-DTW.²⁷ The number of times the k-means algorithm would be run with different centroid seeds was set to "3" and the number of iterations for the barycenter computation process was set to "10." The clustering was carried out repeatedly, with a preset of between 2 and 10 clusters.

The optimal number of clusters was determined by the so-called elbow-method. This method visualizes the sum of squared distances of samples to their closest cluster center (inertia). To determine the optimal number of clusters, one selects the value of k at the "elbow", the point after which the inertia start decreasing in a linear fashion.²⁸

3. Results

3.1 Proportion of books with genre labels

An initial important question may be to ask how many books have genre terms in their titles at all. The first graph (see figure 2, below) shows that there have been considerable changes regarding this question:



Figure 2: Books with genre labels, 1500-2020

While it was not uncommon in the 16th and the first half of the 17th century for German Literature to have genre-indicating subtitles (around 30 % of all titles), this choice became unpopular around 1700, when the percentiles dropped below 20 %. Afterwards, there was a steady rise until 1850. By the 1920s, more than 60 % of literary publication added a genre-indicating subtitle. One possible explanation for the drop around 1700 may be the contemporaneous and intense debates about the entire system of genres that had been established at this time.²⁹ These controversial debates, which were in part related to the reception of Nicolas Boileau's strict *L'Art poétique* (1674), may have caused uncertainty among authors regarding existing genre labels. The resulting reluctance to use genre labels subsides during the 18th century with the gradual establishment of a new system of genres and corresponding labels.

3.2 The most common genre labels

The next natural question is: which are the most common genre terms (a) for the early modern period and (b) for the late modern period, including contemporary history? Graphs 3 and 4 (see below) show the *genre appearances* of the 10 most common non-compound terms, again as a proportion of the total literary production.³⁰ These terms are *Roman* (“novel”), *Erzählung* (“narration”), *Novelle* (“novella”), *Lied* (“song”), *Gedicht* (“poem”), *Historie* or *Historia* (“history”), *Epos* (“epic poem”), *Gebet* (“prayer”) and *Brief* (“letter”). Although not in the top ten, I also manually added the terms *Drama / Schauspiel* (“drama / play”) as a further major non-compound genre.

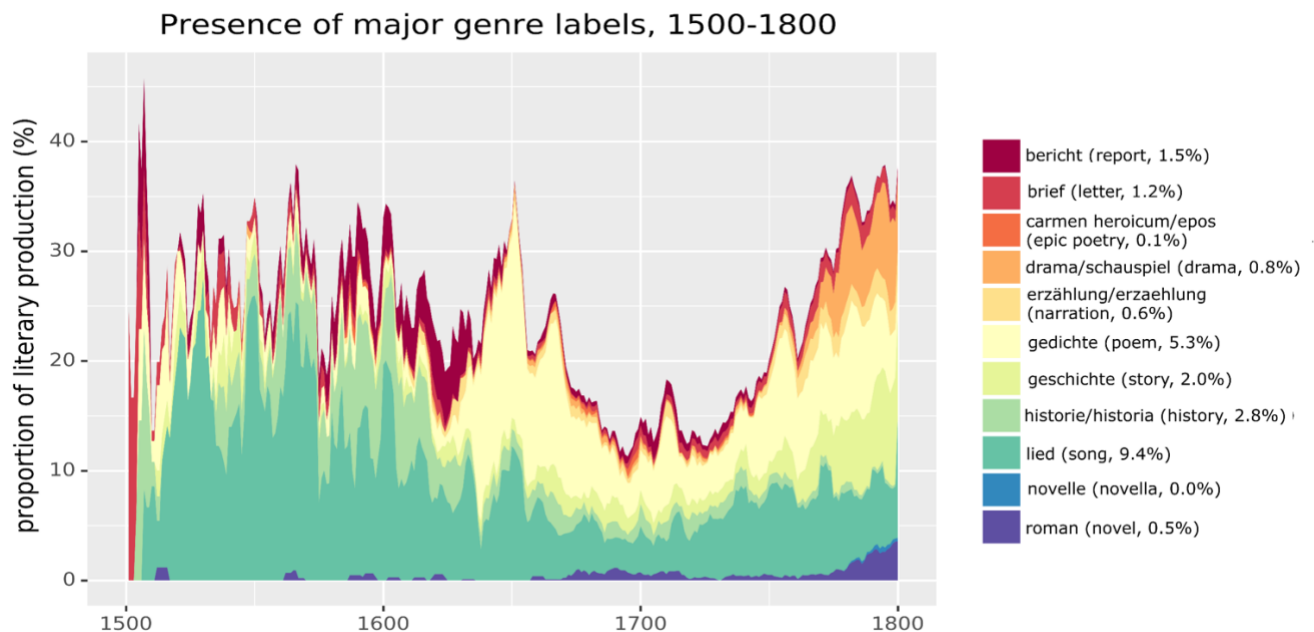


Figure 3: Presence of major genre labels, 1500-1800

As regards the early modern period (see figure 3, above), two genre labels are dominant: the *Lied* (“song”) label from around 1500 until 1650, and the *Gedicht* (“poem”) label from approximately 1625 until 1800. The second finding in particular, however, must be treated with caution: the term *Gedicht* could originally designate all kinds of texts in verse, and only in the last third of the 18th century did the term start to be used for shorter lyric texts as continues through to today.³¹

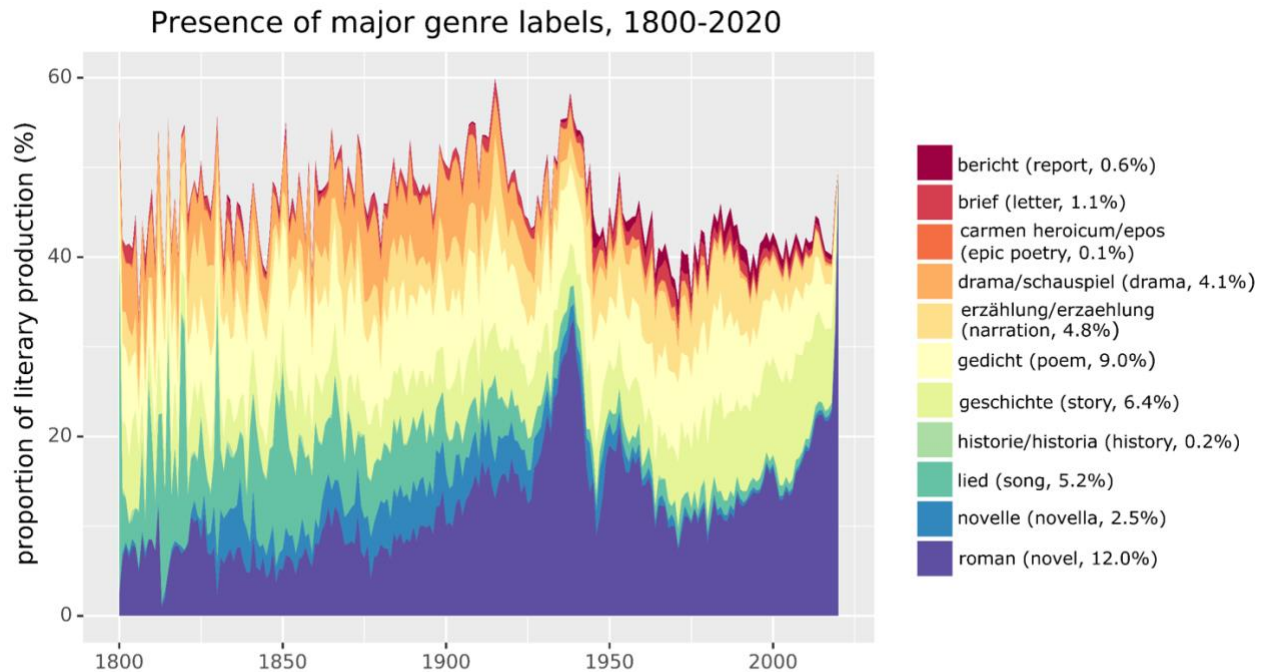


Figure 4: Presence of major genre labels, 1800-2020

Regarding the period from 1800 until today (see figure 4, above), the distribution of generic terms is far more balanced. Seen over the full timespan, there are four significant genre terms: the *Roman* (“novel”), the *Geschichte* (“story”), the *Erzählung* (“narration”) and the *Gedicht* (“poem”). The *Lied* (“song”), which is still important throughout the 19th century, vanishes after 1945. A special case is the genre label *Novelle* (“novella”), which establishes itself only during the first half of the 19th century, but becomes insignificant just one century later during the beginning of the 20th century. The modern period also witnesses the multi-step “rise of the novel” (Ian Watt) as a genre label. The first notable rise is situated in the last third of the 18th century (see figure 3, p. 10), culminating at 1800, when 8% of all literary works have *Roman* in their subtitles. Since the history of the German novel starts earlier than the last third of the 18th century,³² moreover, the rise of the text type clearly predates the rise of the genre term’s usage. A second rise of the genre term’s use takes place during the second half of the 19th and the beginning of the 20th centuries, with a peak in 1939, where 33% of all literary works have the subtitle “Roman”.³³

3.3 Genre terms and genre production

Since there is no reliable data available about the production of German novels,³⁴ the relationship between the use of genre terms and genre production cannot yet be studied for the German novel as a whole. However, there is detailed data concerning the historical novel. A collaborative research project at the University of Innsbruck in the early 2000s manually compiled a list of all historical novels based on the following definition: a historical novel is a fictional work of at least 150 pages, whose plot mainly takes place before the author's birth.³⁵ By applying data from this project, it is possible to relate genre production to labeling (see figure 5, below).

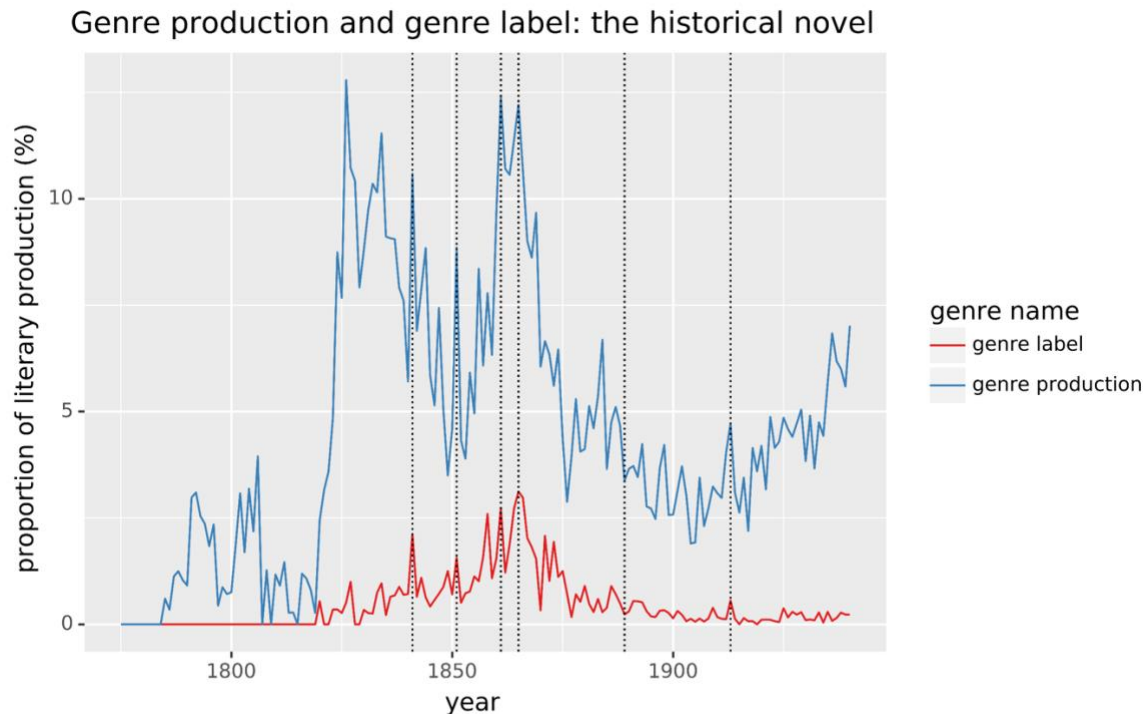


Figure 5: Genre production and genre label, e.g. the historical novel

One can observe at least two things in figure 5. First, this graph shows that the boom of historical novels in the German language, which starts with the translations and pseudo-translations of Walter Scott's novels, is soon followed by a rise in the use of the genre term "historical novel." Second, while the historical novel's first two boom periods, lasting roughly from 1825 until 1850 and 1860 until 1875, are reflected in growing use of the genre label, the further development of the genre is not.

Nevertheless, the continuous use of the genre term fits with the continuous production of historical novels over this period (in contrast to many other discontinuous genre terms). Overall, there is clear correlation between the two data sets, as illustrated by the vertical lines in figure 5 that mark characteristic peaks and valleys. This visual impression is also confirmed by statistical tests. If Spearman's Rho is determined as a suitable correlation measure, a positive correlation is obtained ($r = 0.76$).

This analysis on the example of “historical novels” suggests that the frequency of genre labels is not only, as argued at the beginning of this article, a good indicator for the institutionalization of a genre, but that it also correlates to a certain degree with genre production. Future research will be needed to further illuminate this correlation. In the following sections, I will focus primarily on genre concepts as indicators for the institutionalization or deinstitutionalization of a genre.

3.4 Institutional cycles of genres

It has been argued that generic subtitles are a good indicator of genres' institutionalization, because the usefulness of generic subtitles depends on readers' generic expectations. For major genre terms like “novel,” “history,” or “poem” interpretations are straightforward, with the corresponding genre terms having been clearly established over long periods of time (see figures 3 and 4). However, there are many more genres that never achieve this level of “market share” in the genre label market; they are only institutionalized for short periods. This section considers examples of such “minor” genre labels, working with compound genre terms and phrasal genre terms (adj. + genre).

It's instructive to start with subgenres of the novel, because they are an especially complex and much researched case. The novel's success as the most productive genre in the 19th and 20th century, together with a growing market for literature, created a need for orientation and organization. Authors and publishers sought to fulfil this need through the production of virtually countless genre labels. My analysis shows that for the period between 1800 and today there are approximately 1080 compound genre terms ($x + roman$) and 400 genre terms of the pattern “adjective plus *Roman*” (“novel”).³⁶ The two following figures visualize the 10 most successful of these labels, according to their accumulated proportions of total literary production (see figures 5 and 6, below).³⁷

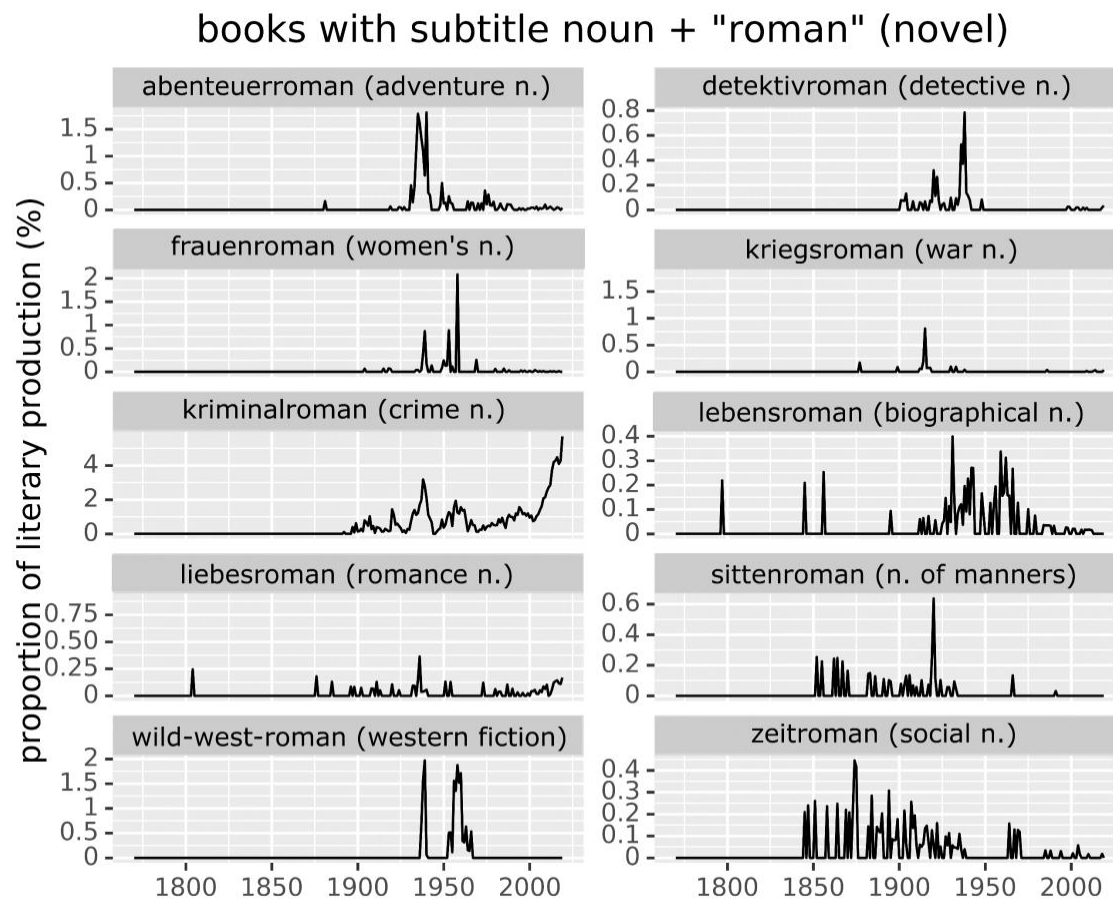


Figure 6: Books with subtitle noun + roman ("novel")

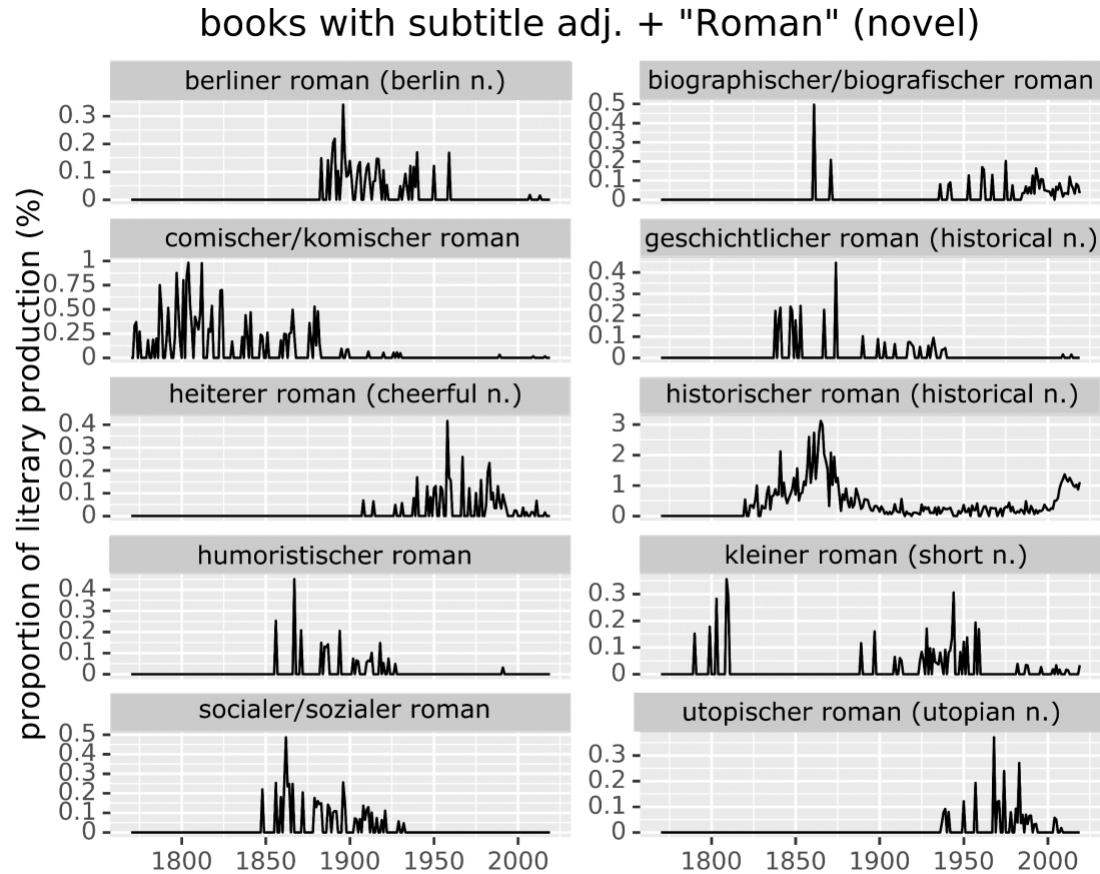


Figure 7: Books with subtitle adjective + Roman ("novel")

There are several distinct distribution patterns. Some graphs indicate a genre's quick rise, like in the case of the war novel (*Kriegsroman*) after World War I or the women's novel (*Frauenroman*). In other cases, however, this process takes several decades (see, e.g., *sozialer/sozialer Roman* or *Berliner Roman*). Almost all of the genre terms show considerable variance over time, with only the romantic novel (*Liebesroman*) existing continuously for more than 150 years without ever becoming particularly popular. In many cases, there is a single clear boom period, exceptions being the small novel (*kleiner Roman*), the biographical novel (*biografischer/biographischer Roman*), and the *Zeitroman* (a special kind of social novel), the distributions of which are discontinuous.

As noted above, figures 6 and 7 represent only the most successful sub-genres of the "novel." This does not include hundreds of less successful novel sub-genres, or the

Even a superficial glance at figure 8 reveals that the vast majority of the most persistent genres are children of the modern age: they are present from the 19th century into the 20th or, frequently enough, the 21st century. Almost all of the genre terms show essentially discontinuous institutional cycles, although an important exception is the historical novel (*historischer Roman*). While shorter interruptions are of little significance, longer interruptions or terminations in genre labels fit surprisingly well with some literary-historical theses about genre developments: the century-long presence of clerical songs (*geistliche Lieder*), the boom of "true stories" in the second half of the 18th century, the institutionalization of the adventure novel (*Abenteuerroman*) during the first third of the 20th century, the establishment of the non-fiction novel (*Tatsachenroman*) and the factual report (*Tatsachenbericht*) in the period of New Objectivity (*Neue Sachlichkeit*), etc. On the other hand, the presence of village stories (*Dorfgeschichten*) in the 20th and 21st centuries is somewhat surprising, as this genre is often seen as a 19th century phenomenon associated with names such as Berthold Auerbach, Gottfried Keller, or Adalbert Stifter.

Only 4 sub-genres of the novel (historical novel, crime novel, adventurous novel, and non-fiction novel) are amongst the 40 most institutionally established genres when selection is made according to the length of their periods of presence. In addition to the novel's classic prose-rivals – narrative (*Erzählung*) and story (*Geschichte*) (cf. figure 4, p. 11) – much less studied genres, such as the "picture" (*Charakterbild*, *Lebensbild*, or *Zeitbild*), or the "report" (*Erlebnisbericht*, *Tatsachenbericht*) also provide rivalry for the novel in this regard.

3.5 Periods of genre differentiation

The moment when authors or publishers start to use a genre term with regularity is of particular interest, because this indicates the moment of the corresponding genre's institutionalization. To determine when the greatest number of genres are at the edge of being institutionalized, I calculated institutional cycles using a slightly looser criterion,⁴⁰ which yielded 581 genres with at least one institutional cycle. The following graph (figure 9, below) represents counts of the beginnings of the first institutional cycle per genre. For example, the local maximum right before 1900 with a y-value of "6" corresponds to six genres that began their first institutional cycle in 1895. In the style of the *previous* graph (figure 8), one would have six bars that begin

in 1895 and there would be no bars before 1895, because these bars represent the *first* institutional cycle of the six genres.

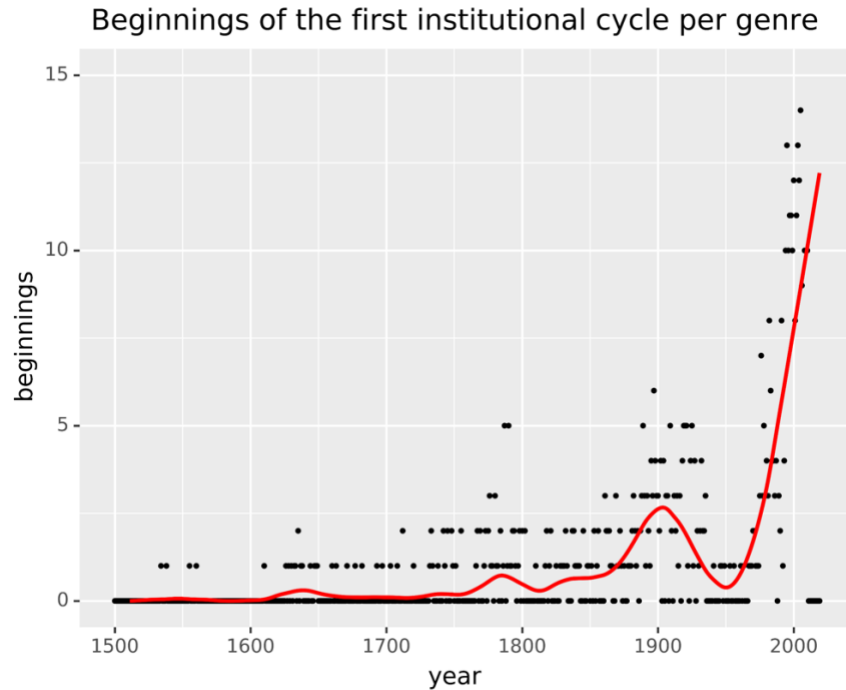


Figure 9: *Beginnings of the first institutional cycle per genre, 1500-2020*

Three periods of genre differentiation can be clearly identified. These are: a first phase from roughly 1750 to 1800, a second from 1875 to 1925, and a third starting in the 1960s. This data is in line with observations made by literary historians. The second half of the 18th century is known as a period in which normative poetics slowly lost control and, according to some theoreticians, the modern system of literature emerged.⁴¹ It would be natural to assume that this change affected authors, too, insofar as they became less interested in complying with “pure genres” (a contemporaneous term) and more willing to experiment with and invent new (sub-)genres.

The second local maximum right after 1900, immediately fits with the observations made by literary historians. Helmuth Kiesel, for example, a renowned specialist on German Literature in the classical modern period, speaks of this period as having a “tendency to dissolve boundaries and overlap or cross literary genres.”⁴² The data in

figure 9 confirm this claim and allow for precision in determining the time periods in which generic differentiation occurs. The particularly sharp increase after the Second World War may be due to a tendency to use genre labels to signal innovation (where such may or may not exist): e.g. the “wine crime novel” (*Weinkrimi*). In general, there is also a tendency in this period to localize established genres, especially with reference to the crime novel, with terms like *Schwarzwaldkrimi*, *Nordseekrimi*, or *Stuttgarkrimi*.

3.6 Patterns of generic subtitles over time

In order to identify patterns across the many time series, I used Dynamic Time Warping (DTW). DTW is a technique to dynamically compare time series data, even if the different time series are offset to each other. The rationale behind using DTW is to stretch or compress time series locally in order to make them resemble one another as much as possible. Using the distance metric of DTW, it is then possible to apply various kinds of clustering techniques to time series data, which identifies groups of time series that resemble each other. The objective is to find similar shapes within a set of time series.

In the present case, the popular KMeans-Clustering algorithm was used on 279 genres (compound genre terms and phrasal terms), which had been previously filtered according to the criterion introduced in section 3.4: genres had to be continuously present for at least 10 years (for details, see section 2.3). Using the so-called elbow-method as heuristics, 3 clusters of time series emerged from the KMeans-clustering (see figure 10, below).

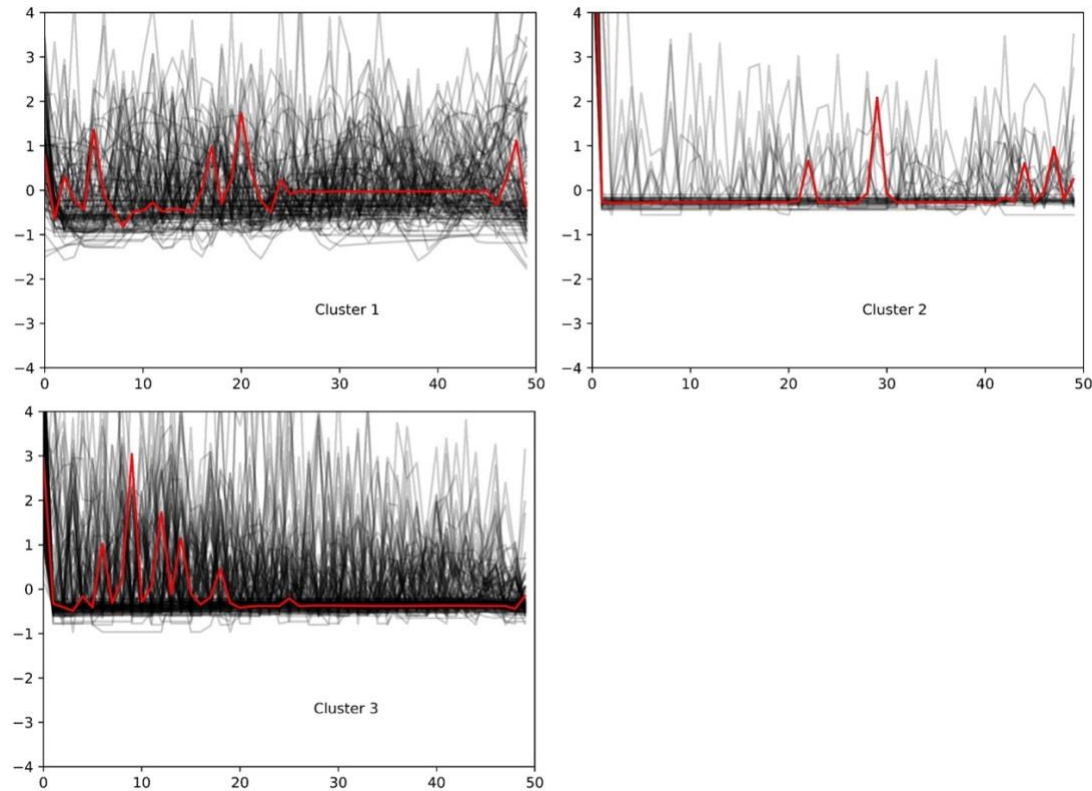


Figure 10: *k*-Means Clusters using Dynamic Time Warping (279 genres)

The red line presents the centroid of each cluster. The lines in grey present the 279 time series, each corresponding to one genre label. It is worth noting that due to the transformation and normalization of the data (see section 2.4 for detail), there is no intuitive interpretation for figure 10's axis *scales*. The x-axis represents time and the y-axis proportion of literary production, but it is not possible to tell from the graph in which concrete year, for example, a rise starts.

Using information about the genres that fall into each cluster (see Appendix A2), it is possible to interpret the clusters as follows. The first cluster is comprised largely of genres that are important over a longer period of time and that experience repeated rises. These include the historical novel, the science fiction novel, the travelogue, and the folk song.⁴³ The second cluster contains discontinuous genres, such as the romance novel, where no clear phase of institutionalization is discernible (see figure 6, p. 14 for comparison). This cluster also includes the biographical novel, erotic poems, satirical poems, the *Spukgeschichte* ("ghost story"), and the travel diary. The

third cluster includes genres that show a clear boom phase, but which have not been permanently institutionalized. These include the edification book, the family novel, the *Sinngedicht* (“epigram”), the social novel, and the *Zeitroman* (special kind of social novel). Remarkably, many genres that emerged in the 19th century can be found in this cluster. More generally, and in a very simplified way, one could refer to the three clusters as single-period-boomers or annuals, evergreens, and permanently institutionalized genres or perennials.

3.7 History-related genres and true stories

Working with the genre data employed in this article, it quickly becomes clear that there are many genre terms containing historical reference. Besides the historical novel, these are genre terms like “historical-romantic play” (*historisch-romantisches Schauspiel*), “historical novella” (*Geschichtsnovelle*) or simply *Historie*. (Although it is well established that in the early modern period the terms *historie* and *historia* are also used for purely invented stories, it remains legitimate to include them here because they may still have *signaled* a relation to history, even if only in the sense of *historia magistra vitae*.⁴⁴) Figure 11, below, depicts the distribution of these terms over the entire period under review.

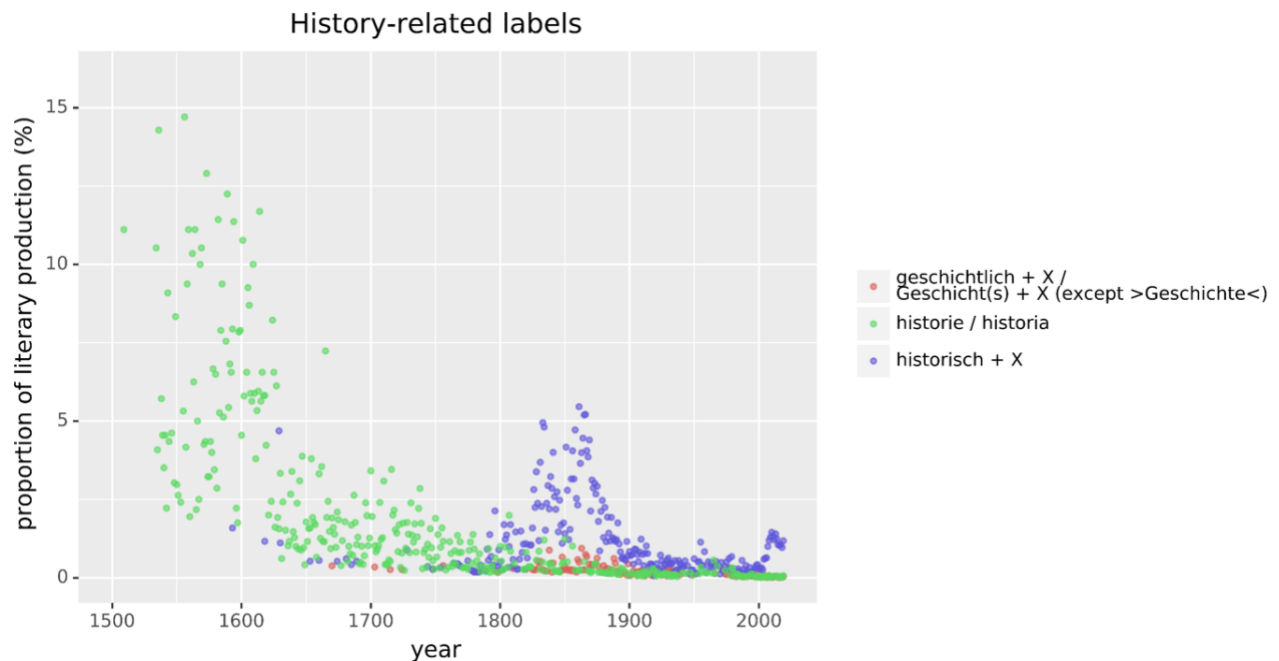


Figure 11: History-related labels, 1500-2020

Figure 11 appears to show a bifurcation in the investigation period. The slow loss of importance of *Historie* and *Historia* (green), which begins in the 18th century, is accompanied by the rise of the adjective *historisch* ("historical", in purple), which plays a dominant role in the 19th century. Only of marginal importance are genre terms with the German equivalent of the Latin loanword "historisch" (*geschichtlich* in German), and of even lesser importance are compositions such as *Geschichtsroman* (aggregated together in the figure in one category, shown in red). The division between the labels "history" and "historical" may mirror two different modes of dealing with history: The early modern period would be the period of (putative) "true histories" (this happened) and the 19th century the time of historical perspective (things used to be thus), which portrays historical periods rather than specific events or persons.⁴⁵

A last observation regarding figure 11: Overall, it is striking that the 18th century, and especially the second half, is characterized by a marginalization of history-related labels. In research, the 18th century has also been known as a period of pseudo-factuality, i.e. a period in which invented stories present themselves as true stories through titles, forewords, or other paratextual framing.⁴⁶ Such framings should also be reflected, at least partially, in works' titles and subtitles (see figure 12, below).

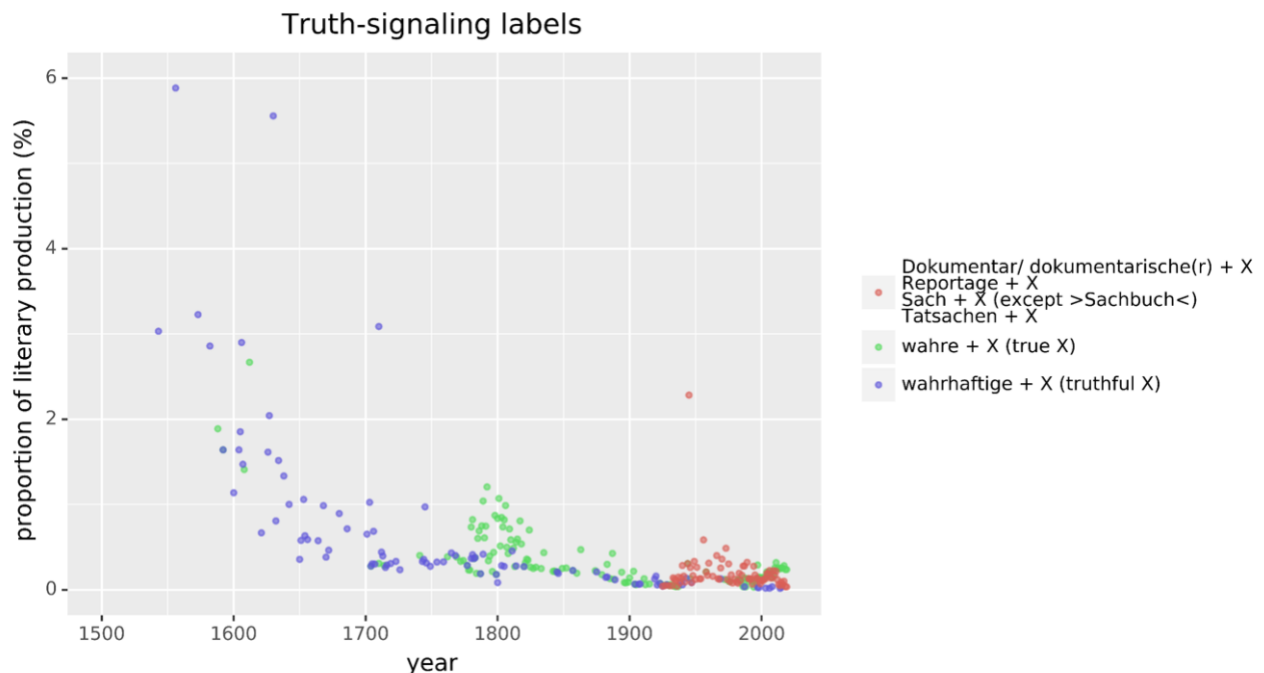


Figure 12: Truth-signaling labels, 1500-2020

Figure 12 shows a roughly three-way division in truth-signaling labels. While the adjective "truthful" (*wahrhaftig*) is continuously present (with different spellings) from 1600 onwards and dominates in the 17th century, two other categories are present in clearly defined periods. The adjective "true" (*wahre*) occurs from the middle of the 18th century; in the 20th century, a group of new compound labels arrive that also signal truth.⁴⁷ The transition from "truthful" to "true" is especially notable, since "truthful" primarily describes a person's characteristics, describing him or her as sincere.⁴⁸ In contrast to "true," a property primarily of propositions or claims, the use of "truthful" in compositions like "truthful story" is metaphoric and leaves room for interpretation. A "true" story thus ought to be considered as the clearest choice for a pseudofactual publication.

Against the background of early modern debates about the relationship between poetry and history (which in turn go back to Aristotle's "Poetics"), it would be natural to ask whether there was a relationship between the development of history-related and truth-signaling genre labels. For the purposes of comparison between the two sets of labels, the categories from figure 11 were combined into one category – "history-related labels" – and the categories from figure 12 into one category – "truth-signaling labels" (see figure 13, below).

The parting of the ways: history-related and truth-signaling labels

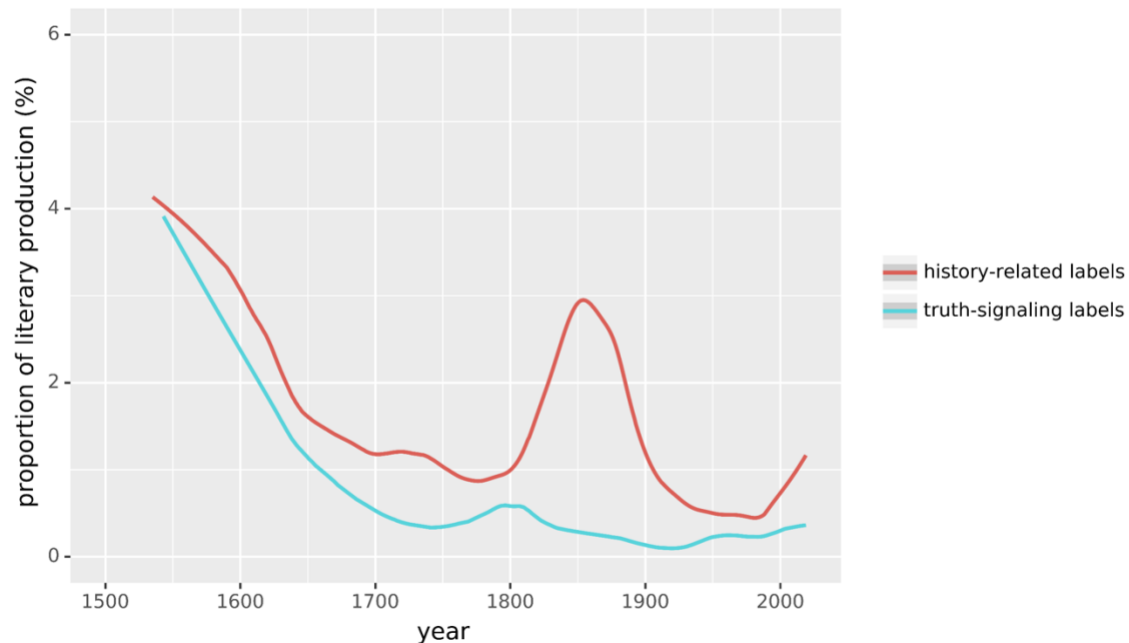


Figure 13: The parting of the ways: history-related and truth-signaling labels, 1500-2020

According to the data here, there is a lengthy parallel development of historically related genre concepts, on the one hand, and truth-claiming genres, on the other. This period of shared development stretches from about 1550 to 1800. At this point, however, a dissociation between the two genre categories occurs: the number of history-related genre labels rapidly increases, while truth-signaling labels decrease. This indicates that increased interest in literary representations of history, which is characteristic for the 19th century (also see figure 5, p. 12 on the historical novel), is *not* accompanied by an increase in truth claims.

At least *prima facie*, these findings seem to fit well with the macro-narrative that Nicholas Paige has developed on the example of French literature. According to Paige, there are three “regimes” of poetic invention. The early-modern Aristotelian regime is characterized by the poet adding “his inventions to the renowned heroes and events of history so as to make a good plot.”⁴⁹ This would correspond to the period of dominance of the term “truthful” (see figure 12, p. 22), since one may *sincerely* (to the best of one's knowledge and belief) fill up and decorate a known (factual) plot.⁵⁰ The intermediate regime is the “pseudofactual” of the 18th century, probably here represented somewhat distorted by the rise of the term “true” (*wahr*, see figure 12), which may seem too late even in the German context. It is worth remembering, however, that a) pseudo-factuality is often signaled in other paratexts, especially prefaces, too and b) that the ‘truth language’ in titles may succeed the rise of production of pseudofactual novels in a similar way the graph on the historical novel (see figure 5, p. 12) illustrates.⁵¹ Finally, the third modern regime in Paige's structure is characterized by a newly implicit agreement to “accept the writer's inventions as a kind of model of reality.”⁵² Under this agreement, renewed interest in history and literary representations can be dissociated from truth-signaling labels (as per figure 13, p. 23).

Conclusion

As this article demonstrates, the long-term development of generic subtitles is a complex and important subject of investigation. Genre terms provide more than just an important orientation function for readers, making them an important indicator of a genre's institutionalization. As the comparison of the production of the historical

novel and the use of the genre label “historical novel” suggests, moreover, there is a strong correlation between genre productivity and the relative frequency of genre subtitles.

The proportion of works with genre-indicating subtitles varies considerably over the period under study, from under 20% of all literary production around 1700 to over 60% shortly after 1900. The most common non-compound genre terms are *Roman* (“novel”), *Erzählung* (“narration”), *Novelle* (“novella”), *Lied* (“song”), *Gedicht* (“poem”), *Historie* or *Historia* (“history”), *epos* (“epic poem”), *Gebet* (“prayer”), and *Brief* (“letter”). If to summarize, the song may be said to dominate the 16th century, the poem the 17th, and the novel the 20th century. The 18th and the 19th century have no clearly dominant genre label. Institutional cycles for the 20 most important novel genres and the 40 most important genres in general have also been determined (see figures 6, 7, and 8). Analysis of the institutionalization phases of 581 sub-genres, moreover, shows that there are two major phases of genre differentiation: first, around 1780, and second, around 1900.

Furthermore, comparative analysis of history-related genre terms, on the one hand, and truth-signaling labels, on the other, supports existing theses in the field of literary history. According to these arguments, the relationship between literature and truth fundamentally changed during the 18th century. Whereas in the early modern period interest in history was accompanied by authors’ signaling of literary claims to truth, the linked development of these two types of genre labels fell apart after 1800.

Finally, the clustering procedures detailed in this article were able to show that the development of generic subtitles follows characteristic patterns. Three such patterns – “one-period-boomers,” “evergreens,” and “perennials” – have been identified through the use of K-Means Clustering. Against the backdrop of the underlying relation between a genre’s productivity and the use of corresponding genre terms, the existence of patterns seems logical. Especially in the modern period, market mechanisms are involved in the production of certain text types, and are likely to be hidden behind the patterns seen in genre labels’ development.

Some limitations of the paper should also be pointed out. First, the database of German Literature created is based on a modern (librarians’) concept of literature that did not exist (as explicit concept) before the 18th century. This problem which

has been intensely debated in the research literature,⁵³ could mean, in the worst case, that the analyses above miss certain dynamics of the genre system which manifest in the titles of books that have been excluded as “non-literature” by the librarians. Future research (which, however, must also presuppose a certain concept of literature) will show how big this problem is. A second limitation concerns the accuracy of the filtering process to create the data set. Since very few of the filtered books were available digitally and the percentage of false positives is likely to vary widely by time period (for example due to books about literature falsely tagged as literature), a quantitative evaluation of the filtering process is a quite challenging task that could not be solved within the framework of a one-man project. Third, due to the exclusion of translations from the dataset the presented analyses cannot do justice to the importance of translation of generic terms from Latin, French and English for the historical dynamics of the system of genre. These influences would need to be scrutinized in a separate analysis.

I am confident that my results, despite these limitations, illustrate that a social-institutional perspective on literary history is not only possible with the help of quantitative methods, but that it can benefit greatly from quantitative methods. Concepts such as genre awareness, genre expectations, interpretive conventions, and reading practices, which have long been established in literary theory, cry out to be operationalized and applied to historical periods and particular communities of historical actors because they all focus on regularities, although of various kinds. Complementing the DH-based research on text types as generalizations over textual features by research on *genre* understood as generalization over individuals’ attitudes will not only have the desirable effect of fostering further exchange between the digital humanities and the literary studies community. Ultimately, on the horizon becomes apparent the possibility of examining how genres as social-institutional entities relate to feature-based text types.

Appendices

A1: Basic genre terms

#	term	#	term
1	Aphorismus	41	Hymne/Hymnus
2	Aufzeichnung	42	Interview
3	Ballade	43	Komödie
4	Bericht	44	Krimi
5	Bild	45	Legende
6	Biografie	46	Lied
7	Brief	47	Lyricum
8	Buch	48	Lyrik
9	Carmen heroicum	49	Märchen
10	Chronik	50	Monolog
11	Darstellung	51	Mythos
12	Dialog	52	Novelle
13	Dichtung	53	Ode
14	Dokumentation	54	Parabel
15	Drama	55	Parodie
16	Drama	56	Poesie
17	Einakter	57	Prosa
18	Ekloge	58	Reportage
19	Elegie	59	Robinsonade
20	Epigramm	60	Roman
21	Episode	61	Romanze
22	Epistel	62	Saga
23	Epos	63	Satire
24	Erzählung	64	Schauspiel
25	Essay	65	Schrift
26	Fabel	66	Skizze
27	Farce	67	Sonett
28	Feature	68	Sylva
29	Feuilleton	69	Szene
30	Fiktion	70	Tragödie
31	Fragment	71	Trauerspiel
32	Gebet	72	Utopie
33	Gedicht	73	Witz
34	Gemälde		
35	Geschichte		
36	Gleichnis		
37	Groteske		
38	Historia		
39	Historie		

A2: Most persistent genre terms and cluster assignment (DTW)

#	genre label	cluster
1	5-minutengeschichte	1
2	abenteuergeschichte	1
3	adventsgeschichte	1
4	alpenkrimi	1
5	autobiografie/autobiographie	1
6	baden-württembergkrimi	1
7	bildergeschichte	1
8	bildermausgeschichte	1
9	biographie	1
10	bodenseekrimi	1
11	bremenkrimi	1
12	bürgerliches trauerspiel	1
13	detektivroman	1
14	dorfgeschichte	1
15	drehbuch	1
16	eifelkrimi	1
17	erlebnisbericht	1
18	erotische geschichten	1
19	erotischer roman	1
20	erzählprosa	1
21	familiensaga	1
22	fantasyroman	1
23	fußballgeschichte	1
24	fußballwitz	1
25	geistergeschichte	1
26	geschichtenbuch	1
27	gruselgeschichte	1
28	gute-nachtgeschichte	1
29	gutenachtgeschichte	1
30	hamburgkrimi	1
31	heimatroman	1
32	historische novelle	1
33	historischer roman	1
34	historisches schauspiel	1
35	hohenlohekrimi	1
36	inselgeschichte	1
37	inselkrimi	1

38	jagdgeschichte	1
39	kaffeehauskrimi	1
40	katzengeschichte	1
41	kinderbuch	1
42	kleine geschichte	1
43	kleine prosa	1
44	kleine romane	1
45	kleinstadtgeschichte	1
46	krimigeschichte	1
47	kriminalgeschichte	1
48	kriminalroman	1
49	kurzgeschichte	1
50	kurzprosa	1
51	kurzroman	1
52	küstenkrimi	1
53	lebensbericht	1
54	lebensbild	1
55	lesebuch	1
56	meistererzählung	1
57	minutengeschichte	1
58	mundartdichtung	1
59	mundartgedicht	1
60	mutmachgeschichte	1
61	mysteryroman	1
62	münsterlandkrimi	1
63	nordseekrimi	1
64	originalroman (aggreg.)	1
65	persönlicher bericht	1
66	phantastische erzählungen	1
67	phantastische geschichten	1
68	provinzkrimi	1
69	reisebericht	1
70	reiseerzählung	1
71	reiseroman	1
72	romanbiografie	1
73	romantische erzählung	1
74	rügenkrimi	1
75	satirische geschichten	1
76	schwabenkrimi	1
77	schwarzwaldkrimi	1
78	schwäbische geschichten	1
79	schülerwitz	1
80	science-fictionroman	1
81	sportgeschichte	1

82	stuttgartkrimi	1
83	syltkrimi	1
84	tagebuch	1
85	tagebuchaufzeichnung	1
86	taschenbuch	1
87	tatsachenbericht	1
88	tatsachenroman	1
89	thüringenkrimi	1
90	tiergeschichte	1
91	utopischer roman	1
92	vaterländisches schauspiel	1
93	volkslied	1
94	volksmärchen	1
95	vorlesebuch	1
96	vorlesegeschichte	1
97	vortragbuch	1
98	wahren geschichte	1
99	weihnachtsgeschichte	1
100	weihnachtskrimi	1
101	weinkrimi	1
102	weisheitsgeschichte	1
103	westernroman	1
104	wissenschaftlich-phantastischer roman	1
105	zeitbild	1
106	zeitzeugenbericht	1
107	abentheuerliche geschichte	2
108	bilderbuch	2
109	biographischer/biografischer roman	2
110	denckwürdige historien	2
111	erotische gedichte	2
112	gelegenheitsgedicht	2
113	gesangbuch	2
114	glückwunschgedicht	2
115	hamburgische geschichte	2
116	heldengeschichte	2
117	kinderlied	2
118	klaggedicht	2
119	kriminalgroteske	2
120	kriminalkomödie	2
121	kurze geschichte	2
122	kurze geschichten	2
123	kurzgedicht	2
124	künstlerroman	2
125	liebesbrief	2

126	liebeslied	2
127	liebesroman	2
128	lobgedicht	2
129	lustige geschichten	2
130	mercklichen geschichten	2
131	phantastischer roman	2
132	reisegeschichte	2
133	reisetagebuch	2
134	satirische gedichte	2
135	schattenbild	2
136	schwäbische gedichte	2
137	schäfergedicht	2
138	spukgeschichte	2
139	tafellied	2
140	vorbericht	2
141	wanderlied	2
142	weihnachtsgedicht	2
143	zeitgeschichte	2
144	abenteuerroman	3
145	abschiedsgedicht	3
146	alltagsgeschichte	3
147	autobiographischer/autobiografischer roman	3
148	bauernkomödie	3
149	bauernroman	3
150	bergroman	3
151	berliner roman	3
152	bibelgeschichte	3
153	bilderbuchgeschichte	3
154	bildungsroman	3
155	bärengeschichte	3
156	characterbild/charakterbild	3
157	comischer/komischer roman	3
158	criminalgeschichte	3
159	culturbild	3
160	detektivgeschichte	3
161	dichterbuch	3
162	dramatische dichtung	3
163	dramatisches gedicht	3
164	ehrengedicht	3
165	epische dichtung	3
166	episches gedicht	3
167	erbauungsbuch	3
168	erfahrungsbericht	3

169	erlebte geschichten	3
170	familiengemählde	3
171	familiengemälde	3
172	familiengeschichte	3
173	familienroman	3
174	freundschaftsgeschichte	3
175	gebetbuch	3
176	geistesgeschichte	3
177	geistliche(r) lieder	3
178	genrebild	3
179	hausbuch	3
180	heimatgeschichte	3
181	heitere gedichte	3
182	heitere geschichten	3
183	heiterer roman	3
184	heldengedicht	3
185	historisch-romantisches gemälde	3
186	historische erzählung	3
187	historischer erzählung	3
188	historischer schauspiel	3
189	historisches trauerspiel	3
190	hochzeitgedicht	3
191	hochzeitsgedicht	3
192	horrorgeschichte	3
193	hundegeschichte	3
194	jugendbuch	3
195	jugendgedicht	3
196	jugendroman	3
197	kalendergeschichte	3
198	kindergebet	3
199	kindergedicht	3
200	kindergeschichte	3
201	kinderwitz	3
202	kirchenlied	3
203	kleine schriften	3
204	kleiner roman	3
205	klostergeschichte	3
206	kriegstagebuch	3
207	kriminalerzählung	3
208	kriminalnovelle	3
209	kurze prosa	3
210	lebensgeschichte	3
211	lebensroman	3

212	liebesgedicht	3
213	liebesgeschichte	3
214	liederbuch	3
215	litteraturgeschichte	3
216	logbuch	3
217	lyrische gedichte	3
218	lyrische prosa	3
219	lügengeschichte	3
220	mordgeschichte	3
221	märchen	3
222	märchenbuch	3
223	märchendrama	3
224	nachdichtung	3
225	naturlyrik	3
226	novellenbuch	3
227	ostergeschichte	3
228	pferdegeschichte	3
229	piratengeschichte	3
230	politische lyrik	3
231	prosagedicht	3
232	reisebild	3
233	rittergeschichte	3
234	romanbiographie	3
235	romantisches gemälde	3
236	romantisches schauspiel	3
237	sachbuch	3
238	sagenbuch	3
239	schulgeschichte	3
240	sensationsroman	3
241	singgedicht	3
242	sinnbild	3
243	sinngedicht	3
244	sittenbild	3
245	sittengemälde	3
246	sittenroman	3
247	skizzenbuch	3
248	socialer/sozialer roman	3
249	sommergeschichte	3
250	soufflierbuch	3
251	soziales drama	3
252	spannende geschichten	3
253	spiegelbild	3
254	stadtgeschichte	3

255	stimmungsbild	3
256	theatergeschichte	3
257	tiergedicht	3
258	tragikomödie	3
259	trauergedicht	3
260	trauerlied	3
261	trostgedicht	3
262	trostlied	3
263	unglaubliche geschichte	3
264	volksbuch	3
265	volkserzählung	3
266	volksschauspiel	3
267	wahre geschichte/wahre geschichten	3
268	weihnachtsbuch	3
269	weihnachtserzählung	3
270	weihnachtslied	3
271	weihnachtsmärchen	3
272	weltliche lieder	3
273	wild-west-roman (aggreg.)	3
274	wintergeschichte	3
275	zaubermärchen	3
276	zeitgedicht	3
277	zeitroman	3
278	zeitschrift	3
279	zukunftsroman	3

References

¹ I sincerely thank Nicholas Paige (Berkeley), Andrew Piper (Montreal), Caroline Sporleder (Göttingen), Thomas Wescott (Göttingen), Dirk Werle (Heidelberg), and my anonymous reviewers for helpful comments and advice on previous drafts of this paper. I also sincerely thank the D-Lab community at the University of California, Berkeley, especially Christopher Hench, for their support in the initial phase of the project from which this paper emerged. The project was funded by the German Academic Exchange Service (DAAD) and the German Research Foundation (DFG) - project number 392946741. Finally, my sincere thanks go to various employees of the Library Network GBV and the German National Library (DNB), who provided invaluable information to carry out this project, especially to Andrea Hemmer (DNB), Reinhold Heuvelmann (DNB), Bettina Kunz (GBV), Michael Rzehak (GBV), and Gerald Steilen (GBV).

² For this distinction see Harald Fricke, *Norm und Abweichung. Eine Philosophie der Literatur* (München: Beck, 1981), 130–33, as well as more recently Benjamin Gittel, *Fiktion und Genre. Theorie und Geschichte referenzialisierender Lektürepraktiken 1870–1910* (Berlin, Boston: De Gruyter 2021) [forthcoming], ch. I.6. Also see the related distinction between “genre” and “genre membership” in Gregory Currie, *Arts and Minds* (Oxford, New York: Oxford University Press, 2004), ch. 3.

³ Sarah Allison, Ryan Heuser, Matthew Jockers, Franco Moretti, Michael Witmore, “Quantitative Formalism: An Experiment.” Stanford Literary Lab, <https://litlab.stanford.edu/LiteraryLabPamphlet1.pdf>; Mariona Coll Ardanuy and Caroline Sporleder, “Structure-based Clustering of Novels.” Proceedings of the 3rd Workshop on Computational Linguistics for Literature, CLfL EACL 2014: 31–39. DOI: 10.3115/v1/W14-0905; Lena Hettinger et al., “Classification of Literary Subgenres (DHD-Conference Paper).” <http://dhd2016.de/abstracts/votr%c3%a4ge-049.html> (accessed August 12, 2020); Andrew Piper, “Fictionality,” *Journal of Cultural Analytics* 1, no. 1 (2017). DOI:10.22148/16.011; Christof Schöch, “Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama,” *Digital Humanities Quarterly* 11, no. 2 (2017); Ted Underwood, “The Life Cycles of Genres,” *Journal of Cultural Analytics* 1, no. 1 (2017) DOI: 10.22148/16.005; Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change* (Chicago: The University of Chicago Press, 2019) DOI: 10.7208/chicago/9780226612973.001.0001.

⁴ See Piper, “Fictionality,” esp. 9–12.

⁵ See Underwood, *Distant Horizons*, 41–44.

⁶ See Wilhelm Voßkamp, “Gattungen als literarisch-soziale Institutionen. Zu Problemen sozial- und funktionsgeschichtlich orientierter Gattungstheorie und -historie,” in *Textsortenlehre, Gattungsgeschichte*, eds. Alexander Bormann and Walter Hinck (Heidelberg: Quelle und Meyer, 1977), 27–42.

⁷ Cf. the related approaches in Matthew L. Jockers, *Macroanalysis. Digital Methods and Literary History* (Urbana: University of Illinois Press, 2013), 57–62, Franco Moretti, “Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740–1850).” *Critical Inquiry* 36 (1), 2009: 153–58. DOI: 10.1086/606125 and even the early work of Marianne Spiegel, *Der Roman und sein Publikum im früheren 18. Jahrhundert: 1700–1767* (Bonn: Bouvier, 1967). Spiegel postulates “strict conventions” concerning titles and subtitles that enabled the 18th century reader of novels to discern only three subgenres of the novel. However, it is not entirely clear how she arrived at her statistics concerning the production of these three genres, the names of which do not match the signaling terms in the titles. See *ibid.*, 7–33, here 10.

⁸ Ted Underwood and the NovelTM Research Group, “Genre Theory and Historicism,” *Journal of Cultural Analytics* 1, no. 1 (2017) DOI: 10.22148/16.008.

⁹ Reflections on the relevance of these three criteria can be found, inter alia, in: Currie, *Arts and Minds*, ch. 3; Fricke, *Norm und Abweichung*, ch. 4.1; Peter J. Rabinowitz, “The Turn of the Glass Key: Popular Fiction as Reading Strategy,” *Critical Inquiry* 11, no. 3 (1985); Voßkamp, *Gattungen als literarisch-soziale Institutionen*; Kendall L. Walton, “Categories of Art,” *The Philosophical Review* 79, no. 3 (1970).

¹⁰ See Gittel, *Fiktion und Genre*, especially ch. I.6. and II.3, which empirically examines genre-specific reading practices, too.

¹¹ For an examination of the rules of dealing with subgenres of the novel, cf. *ibid.*

¹² On types and possible functions of generic paratexts in general see Gérard Genette, *Paratexts: Thresholds of Interpretation* (Cambridge: Cambridge Univ. Press, 2001), 10–12 and 94–103.

¹³ See the indications in note 7 and note 36, however.

¹⁴ Existing corpora of German Literature, such as the German Text Archive (DTA), the Corpus of Literary Modernism (Kolimo), and the Corpus of German Novels (DROC) are not designed for comprehensiveness and bibliographies of German Literature like Gero v. Wilpert and Adolf Gühring, *Erstausgaben Deutscher Dichtung. Eine Bibliographie zur deutschen Literatur 1600–1960* (Stuttgart, 1967) explicitly exclude “third and fourth rank authors long forgotten and unimportant today” (*ibid.*, Pref.). A separate project on over-length books has created a database from the data of the German National Library, but this project has very different objectives and therefore different filtering criteria: for example its database does not filter for first editions (see Frank Fischer and Ubbo Veenster, “Empirical Data on Over-Length Books,” <https://welitliteratur.net/Empirical-Data-on-Over-Length-Books/> (accessed August 7, 2020)).

¹⁵ See note 1 please for my sincere thanks.

¹⁶ For detail, see the script “Generic_Subtitles_filtering” provided together with the database. To read the data, “MARCRReader” from the package pymarc (version 4.0.0) was used.

¹⁷ See “Language Detection API,” <https://github.com/detectlanguage/detectlanguage-python> (accessed July 27, 2020).

¹⁸ See the header of this paper. GBV’s data set, unfortunately, is not freely available, but can be requested for research purposes.

¹⁹ The data was downloaded on 01/10/2019, but already contained records for 2020 and beyond, probably due to collaborations with larger publishing houses.

²⁰ The analysis always takes title and subtitle into account, because this distinction is not coherent in the GBV-data, e.g. the data field for the title contains what we would call a subtitle, too.

²¹ The list derives from a German handbook for literary genres (see Dieter Lamping et al., eds., *Handbuch der literarischen Gattungen* (Stuttgart: Alfred Kröner Verlag, 2009)) and was supplemented by early modern genres like *Historie* or *Historia* (“history”).

²² See the script “Generic_Subtitles_filtering”. The spaCy-documentation is available online at: <https://spacy.io/models>

²³ See Wes McKinney et al., “Data Structures for Statistical Computing in Python.” Proceedings of the 9th Python in Science Conference, Vol. 445, 2010 , 51–56.

²⁴ See Tavenard, Romain; Faouzi, Johann; Vandewiele, Gilles; Divo, Felix; Androz, Guillaume; Holtz, Chester et al., “Tslern, A Machine Learning Toolkit for Time Series Data.” Journal of Machine Learning Research 21 (118), 2020, 1–6. Available online at <http://jmlr.org/papers/v21/20-091.html>.

²⁵ See Donald J. Berndt and James Clifford, “Using dynamic time warping to find patterns in time series,” AAAI-94 Workshop on Knowledge Discovery in Databases, <https://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-031.pdf>; Eamonn J. Keogh and Michael J. Pazzani, “Scaling up Dynamic Time Warping for Datamining Applications,” in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ed. Raghu Ramakrishnan (New York, NY: ACM, 2000), 285–89.

²⁶ See Hui Ding et al., “Querying and mining of time series data,” *Proceedings of the VLDB Endowment* 1, no. 2 (2008). DOI:10.14778/1454159.1454226.

²⁷ Marco Cuturi, Mathieu Blondel, “Soft-DTW: a Differentiable Loss Function for Time-Series.” ICML 2017. See Purnima Bholowalia and Arvind Kumar, “EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN.” *International Journal of Computer Application* 105, no. 9, 2014.

²⁸ See Purnima Bholowalia and Arvind Kumar, “EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN,” *International Journal of Computer Application* 105, no. 9 (2014).

²⁹ For a survey see Klaus R. Scherpe, *Gattungspoetik im 18. Jahrhundert: Historische Entwicklung von Gottsched bis Herder* (J.B. Metzler, 1968, 18–26).

³⁰ I suppressed *Bild* (“image”) and *Buch* (“book”), which render good results for subgenres like *Tagebuch* or *Zeitbild*, but are not reliable or informative as standalone terms.

³¹ See Rudolf Brandmeyer “Gedicht.” in: *Handbuch der literarischen Gattungen*, ed. Dieter Lamping et al. (Stuttgart: Alfred Kröner Verlag, 2009), 307–15, esp. 307–8.

³² See Spiegel, *Der Roman*; Simons, *Marteaus Europa*.

³³ A remarkable phenomenon, for which no easy explanation is available, is the quite similar increase of publications with the subtitles “novel” (*Roman*) and “poem” (*Gedicht*) in the late 1920 and 1930s, with a peak in 1938.

³⁴ There is data available from the so-called “Codex nundinarius Germaniae literatae bisecularis,” an index of the publishing industry. This index, however, only holds a unitary category for novels from 1801 till 1846, after which point novels are combined together with many other kinds of literature. There are also several well-known corpora (the German Text Archive (DTA), the Corpus of Literary Modernism (KoliMo), and the Corpus of German Novels (DROC)), but none of them is designed to reflect novel production in quantitative terms.

³⁵ See Günter Mühlberger and Kurt Habitzel, “The German Historical Novel from 1780 to 1945: Utilising the Innsbruck Database,” in *Travellers in Time and Space / Reisende durch Zeit und Raum. The German Historical Novel / Der deutschsprachige historische Roman*, eds. Osman Durrani and Julian Preece, (Amsterdam: Rodopi, 2001); and Kurt Habitzel and Günter Mühlberger, “Der historische Roman und sein Beitrag zum Literatursystem der Restaurationszeit (1815-1848/49).” *Internationales Archiv für Sozialgeschichte der deutschen Literatur* 21, no. 1 (1996) DOI: <https://doi.org/10.1515/iasl.1996.21.1.91>. I expressly thank these authors for providing their data for further analysis.

³⁶ Fotis Jannidis has developed an inspiring analysis on the basis of data from google books, compiling a list of approximately 500 subgenres of the novel (see Fotis Jannidis, “Kleiner Versuch über Romangattungen (1. und 2. Teil),” <https://dhd-blog.org/?p=2128>; <https://dhd-blog.org/?p=2384> (accessed August 7, 2020)). Jannidis himself points towards the approach taken here: “The analysis has provided a first impression of the frequency distribution of generic terms, but we cannot judge whether these are terms of the object- or meta-language. The problem could possibly be solved by systematically evaluating the subtitles of novels in a large library catalogue from the perspective of which of the categories listed here occur. This would then prove their use in the object language” (ibid., part 1, my transl.).

³⁷ I excluded the category *Originalroman*, which is kind of pseudo-genre term designating non-translated novels.

³⁸ The seemingly arbitrary number of 10 years is the result of a trial-and-error process. If one chooses a period that is too long, conjunctures of certain subgenres relevant to literary studies disappear; if one chooses a period that is too short, one obtains data that is difficult to interpret because it is too heterogeneous and contains too many genre terms.

³⁹ See section 2.3 for technical detail.

⁴⁰ See section 2.3 for technical detail. Using the 279 genres that evolve from the stricter criterion also produces a peak slightly after 1900, although it makes invisible the local peak at 1780.

⁴¹ See Klaus. R. Scherpe, *Gattungspoetik im 18. Jahrhundert: Historische Entwicklung von Gottsched bis Herder* (J.B. Metzler, 1968), who speaks of a “dissociation of normative genre concepts” (264, my transl.). On the emergence of the “social system of literature,” see Siegfried J. Schmidt, *Die Selbstorganisation des Sozialsystems Literatur im 18. Jahrhundert* (Frankfurt am Main: Suhrkamp, 1989).

⁴² Helmuth Kiesel, *Geschichte der literarischen Moderne: Sprache, Ästhetik, Dichtung im zwanzigsten Jahrhundert* (München: C.H. Beck, 2004), 155.

⁴³ Moreover, there are conspicuously many sub-genres of the crime novel (*Krimi*) in this cluster, which behave similarly in their much shorter existence, compared to the historical novel. It is worth recalling that DTW stretches or compresses the time series locally in order to make one resemble the other as much as possible.

⁴⁴ Knappe claims for the end of the 15th century “a fixed naming convention for prose stories, the events of which are situated in the past and under the influence of historical characters (i.e. above all nobility)” to be called “*Historie*” (Joachim Knappe, “*Historie*” in *Mittelalter und früher Neuzeit. Begriffs- und gattungsgeschichtliche Untersuchungen*

im interdisziplinären Kontext (Baden-Baden: Koerner, 1984), 330, my transl.). Also see Gert Melville „Historie“ in: *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, eds. Harald Fricke, Klaus Grubmüller, Jan-Dirk Müller; Klaus Weimar (Berlin: De Gruyter 1997–2003), vol. 2, 49–52. On the "historia-fabula-argumentum" triad, which was acquired from antique rhetoric for early modern poetological debates, see Stefan Trappen, "Fiktionsvorstellungen der Frühen Neuzeit. Über den Gegensatz von „fabula" und „historia" und seine Bedeutung für die Poetik," *Simpliciana. Schriften der Grimmelshausen-Gesellschaft XX* (1998).

⁴⁵ At least, this dovetails very well with the argument in Barbara Foley, *Telling the Truth: The Theory and Practice of Documentary Fiction* (Ithaca, NY: Cornell University Press, 1986). esp. ch. 5 and 6.

⁴⁶ See Nicholas D. Paige, "Pseudofactuality," in *Narrative Factuality: A Handbook*, ed. Monika Fludernik and Marie-Laure Ryan (Berlin: De Gruyter, 2020), 593–600.

⁴⁷ The aim of this graph is not to visualize non-fictional genres. It would be tempting to model what Monika Fludernik has called "the rise of factuality" (Monika Fludernik, "The Fiction of the Rise of Fictionality," *Poetics Today* 39, no. 1 (2018), 84. DOI: 10.1215/03335372-4265071), but I believe that similar analyses on the basis of my database could be distorted, because librarians (in different time periods) may have used fictionality as an (implicit) criterion of what to categorize as "literature."

⁴⁸ See Jacob Grimm, Wilhelm Grimm, *Deutsches Wörterbuch*. Lfg. 5, 1906, vol. XIII, 1922, col. 825.

⁴⁹ Nicholas D. Paige, *Before Fiction: The Ancien Régime of the Novel* (Philadelphia: University of Pennsylvania Press, 2011), x.

⁵⁰ Such an embellishment of historical plots, which is nevertheless tied to a concept of truth, is much discussed in poetological treatises from the early modern period, see Uwe Korn and Dirk Werle, "Telling the Truth: Fictionality and Epic in Seventeenth-Century German Literature," *Journal of Literary Theory* 14, no. 2 (2020), 241–259.

⁵¹ It is also worth mentioning, that pseudofactual works existed long before the 18th century, although pseudofactual novels took often new forms like the first-person-novel in the 18th century, and that the label "truthful" may have been used for earlier forms of pseudofactuality as well.

⁵² *Ibid.*

⁵³ See Olaf Simons, *Marteaues Europa oder Der Roman, bevor er Literatur wurde: Eine Untersuchung des deutschen und englischen Buchangebots der Jahre 1710 bis 1720* (Amsterdam: Rodopi, 2001), esp. 23–25; Richard Terry, "The Eighteenth-Century Invention of English Literature: A Truism revisited," *Journal for Eighteenth-Century Studies* 19, no. 1 (1996).